High Impact Skills Development Program 

AI & Data Science

Project of Data Mining

**NAME: Nadir Abbas**

**ROLL NO: GIL\_DSAI\_152**

**SECTION: 3rd**

**Email: nadir03128743177@gmail.com**

**Project Title: Online Retail Segmentation**

**Dataset: Retail shop**

**Overview:**

Customer segmentation constitutes a widely adopted strategic approach within organizations, aimed at systematically categorizing clients by analyzing a spectrum of factors such as demographics, shopping patterns, and other discernible traits. By effectively segmenting the customer base, businesses gain the ability to tailor their offerings, communications, and marketing efforts with a higher degree of precision, ultimately leading to enhanced customer satisfaction, targeted engagement, and improved overall business performance.

**Metadata:**

Look for a dataset that contains information about customers such as demographic information, purchasing history, and customer interactions. The data set contains the following variables:

**Invoice No:** The invoice number for each transaction **Stock Code**: The unique code for each product sold

**Description**: The description of each product sold

**Quantity**: The quantity of each product sold in each transaction

**Invoice Date**: The date and time of each transaction

**Unit Price:** The price of each product sold

**Customer ID**: The unique identifier for each customer

**Country**: The country where each transaction occurred

SQL Project Idea: Use SQL queries to answer the following questions:

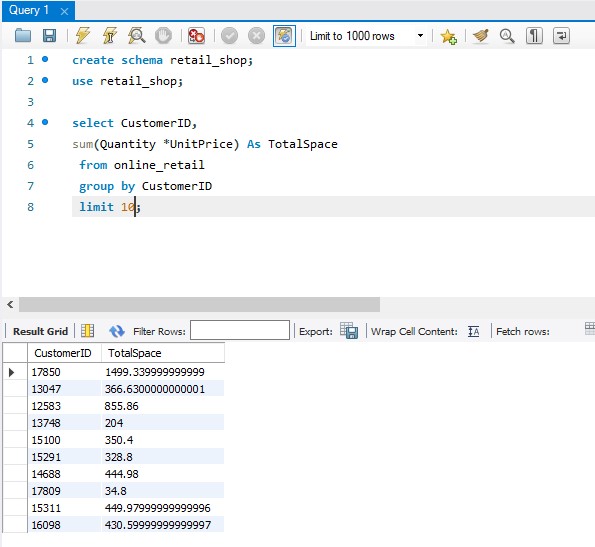
**Beginner Queries.**

Q1: **Define meta data in MySQL workbench or any other SQL tool**

Metadata in SQL tools like MySQL Workbench refers to data that describes other data within the database. This includes:

* **Schema Information**: Details about database structures, such as tables, columns, data types, and constraints.
* **Table Definitions**: Information about table names, column names, data types, and relationships between tables.
* **Index Information**: Data about indexes on tables, including their names and columns they cover.  **User Privileges**: Data about permissions granted to users or roles.

Q2:  **What is the distribution of order values across all customers in the dataset?**

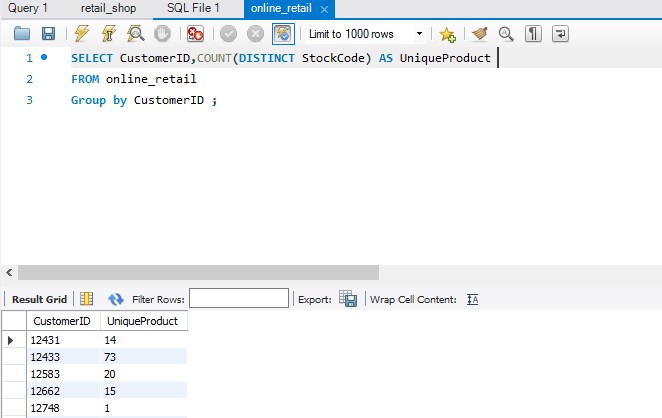


SUMMERY:

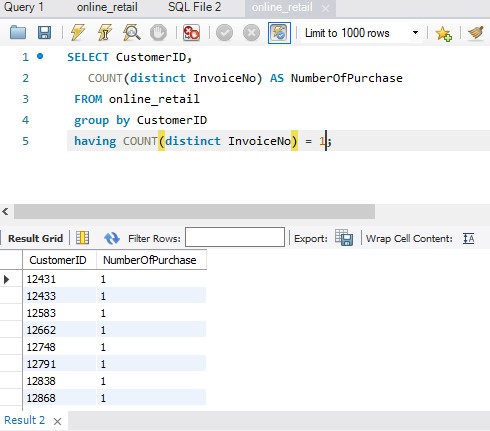
The SQL query calculates the total amount of money each customer has spent on their orders

By summing up the total price of products they have purchased.

**Q3 How many unique products has each customer purchased?**

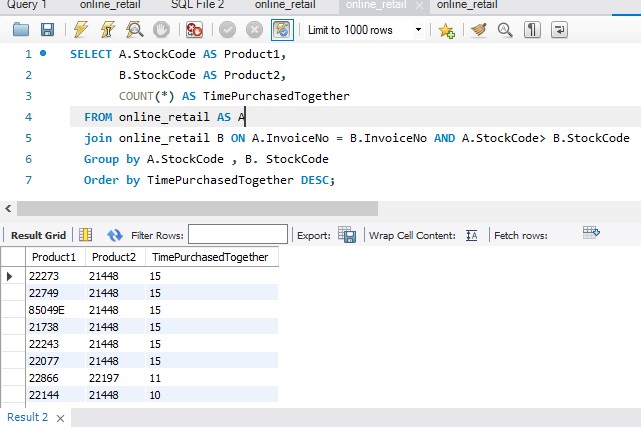


**Q4 Which customers have only made a single purchase from the company?**



This query returns the list of customer ID ‘s for those who have made only single purchase from the company. By using HAVING COUNT(DISTINCT INVOIC NO ) = 1, It only filter out customers with more than purchase ,focusing only on those who have made only single transition .

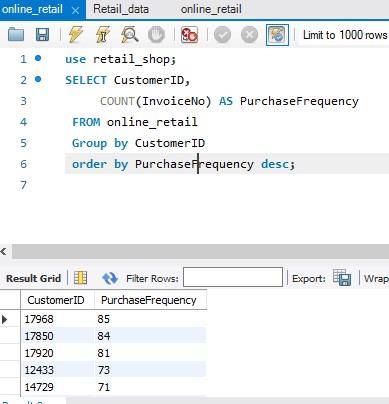
**Q5 Which products are most commonly purchased together by customers in the dataset?**



**Advance Queries**

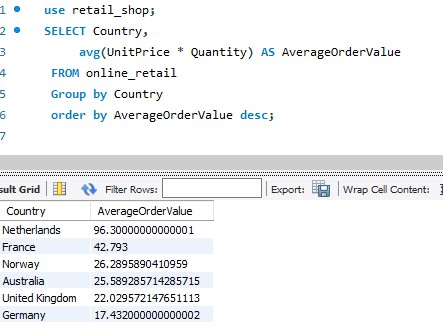
1. **Customer Segmentation by Purchase Frequency**

Group customers into segments based on their purchase frequency, such as high, medium, and low frequency customers. This can help you identify your most loyal customers and those who need more attention.



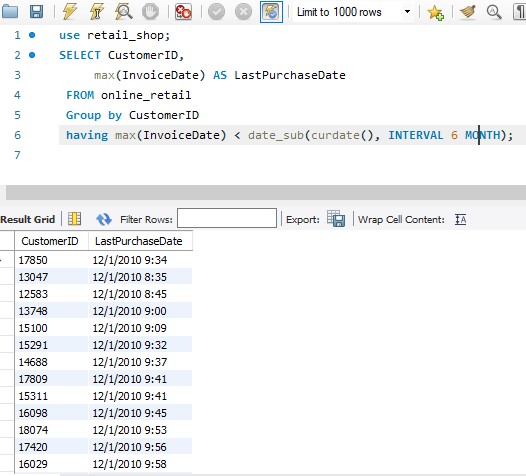
1. **Average Order Value by Country**

Calculate the average order value for each country to identify where your most valuable customers are located.



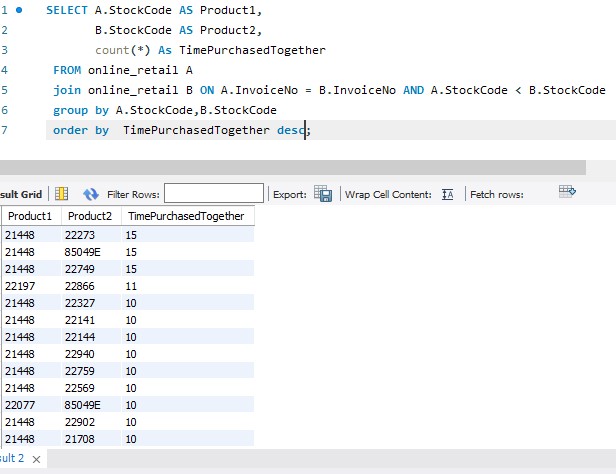
1. **Customer Churn Analysis**

Identify customers who haven't made a purchase in a specific period (e.g., last 6 months) to assess churn.



1. **Product Affinity Analysis**

Determine which products are often purchased together by calculating the correlation between product purchases.



1. **Time-based Analysis**

Explore trends in customer behavior over time, such as monthly or quarterly sales patterns.

